

Detection of additive outliers in Poisson INteger-valued AutoRegressive time series

Maria Eduarda Silva*

Faculdade de Economia, Universidade do Porto,
Rua Dr Roberto Frias, s/n, 4200 464 Porto, Portugal

Isabel Pereira†

Departamento de Matemática, Universidade de Aveiro,
Campus Santiago, 3810 193 Aveiro, Portugal

Center for Research and Development in Mathematics and Applications

December 2011

Abstract

Outlying observations are commonly encountered in the analysis of time series. In this paper the problem of detecting additive outliers in integer-valued time series is considered. We show how Gibbs sampling can be used to detect outlying observations in INAR(1) processes. The methodology proposed is illustrated using examples as well as an observed data set.

Keywords: Additive outliers; Bayesian analysis; Integer-valued time series; INAR(1) model; Gibbs sampler

1 Introduction

This work considers a Bayesian approach to the problem of modelling a Poisson integer valued autoregressive time series contaminated with additive outliers.

It is well known that unusual observations and intervention effects often occur in data sets and can have adverse effects on model identification and parameter estimation. In the framework of Gaussian linear time series the problem of detecting and estimating outliers and other intervention effects has been investigated by several authors including Fox (1972), Tsay (1986), Chang et al. (1988), Chen and Liu (1993) and Justel et al. (2001), among others. However, the problem of modelling outliers and other intervention effects in the context of time series of counts has, as yet, received little attention in the

*Corresponding author: mesilva@fep.up.pt

†isabel.pereira@ua.pt

literature albeit its relevance for inference and diagnostics. Moreover, in this context additional motivation stems from the fact that the usual techniques for outlier removal are not adequate since often lead to non integer values. In the framework of count time series it is worth mentioning the work of Fokianos and Fried (2010) who investigate the problem of modelling intervention effects in INGARCH models and Barczy et al. (2010, 2011) who consider CLS estimation of the parameters of an INAR(1) model contaminated, at known time periods, with innovational and additive outliers, respectively.

The well-known assertion of George Box that while all models are wrong some are useful, motivates that we approach the issue of modelling outliers in integer-valued time series focusing on the integer valued autoregressive model of order 1. In fact, this model introduced independently by Al-Osh and Alzaid (1987) and McKenzie (1985) to model time series of counts, has been extensively studied in the literature and applied to many real-world problems including statistical process control, (Weiß, 2007) because of its simplicity and easiness of interpretation.

To motivate our approach, we represent in figure 1 a data set studied by Weiß (2007) concerning the number of different IP addresses (approximately equivalent to the number of different users) accessing the server of the pages of the Department of Statistics of the University of Würzburg in two-minute periods from 10 am to 6 pm on the 29th November 2005, in a total of 241 observations. This time series is constructed from log data concerning accesses to pages on the server. Weiß (2007) models the data with a Poisson INAR(1) model and using statistical process control techniques finds an outlying observation at time $t = 224$. As described by that author a detailed analysis of the original log data showed that all the eight accesses at that time came from the AOL browser that is known to supply permanently new addresses within a small area. Therefore, it is not possible anymore to infer the user from the IP address. It is interesting to investigate if this observation can be explained by a simple INAR(1) model and if the fit can be improved by the inclusion of an additive outlier effect.

Let $\{X_t\}$ be a Poisson INAR(1) process satisfying

$$X_t = \alpha \circ X_{t-1} + e_t = \sum_{j=1}^{X_{t-1}} \xi_{t,j} + e_t, \quad (1)$$

with $(\xi_{k,j})$ a sequence of Bernoulli r.v. with mean $\alpha \in [0, 1]$ and $\{e_t\}$, the arrival process, a sequence of i.i.d. Poisson variables $e_t \sim \mathcal{Po}(\lambda)$. When additive outliers (AO) occur at times τ_1, \dots, τ_k , with integer sizes $\omega_1, \dots, \omega_k$, X_t is unobservable and it is assumed that the observed series $\{Y_t\}$ satisfies

$$Y_t = X_t + \sum_{i=1}^k I_{t,\tau_i} \omega_i,$$

where $k \in \mathbb{N}$ is the number of outliers and $I_{t,s}$ is an indicator function taking the value 1 if $t = s$ and 0 otherwise. Roughly speaking an additive outlier can be interpreted as a measurement error or as an impulse due to some unspecified exogenous source at time τ_i , $i = 1, \dots, k$.

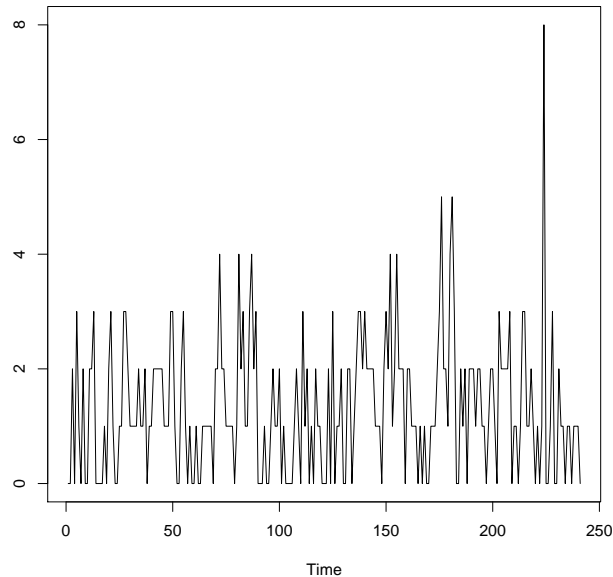


Figure 1: Number of different IP addresses accessing the server of the pages of the Department of Statistics of the University of Würzburg between 10 am and 6 pm on 29 November 2005

Here we consider a Bayesian approach to the problem of Poisson INAR(1) model specification in the presence of additive outliers. Gibbs sampling provides estimates for the probability of outlier occurrence at each time point leading to an effective outlier detection and accurate parameter estimation. Bayesian approaches have been used to detect outliers in ARMA models by Justel et al. (2001) and in bilinear models by Chen (1997).

The paper is organized as follows. Section 2 describes the setup of additive outliers in INAR(1) models and explains the procedure for outlier detection. Section 3 illustrates the methodology on several sets of simulated data as well as on a data set concerning the number of different IP addresses accessing the server of the pages of the Department of Statistics of the University of Würzburg. Section 4 concludes the paper.

2 INAR(1) models with additive outliers

Assume that the observed time series Y_1, \dots, Y_n is generated by

$$Y_t = X_t + \eta_t \delta_t, \quad 1 \leq t \leq n \quad (2)$$

where X_t is a Poisson INAR(1) process satisfying (1), $\delta_1, \dots, \delta_n$ are independent and identically distributed Bernoulli variables with $P(\delta_t = 1) = \epsilon$ and η_1, \dots, η_n are independent random variables identically distributed as $Po(\beta)$. Also, δ_t and η_t are independent for all t . This means that if $\delta_t = 1$ the observation Y_t is contaminated with an AO of magnitude η_t . Note that an outlier at time t affects the model only at instants t and $t + 1$.

2.1 Estimation procedure

In this section we describe the Bayesian approach via Gibbs sampling to estimate model (2). Assume that $Y_1 = X_1$, that is, there is no outlier in the first observation and let $\mathbf{Y} = (Y_1, \dots, Y_n)$, $\boldsymbol{\Theta} = (\alpha, \lambda)$, $\boldsymbol{\delta} = (\delta_1, \dots, \delta_n)$, $\boldsymbol{\eta} = (\eta_1, \dots, \eta_n)$. Now we need to derive the conditional posterior distributions of $\boldsymbol{\Theta}$, $\boldsymbol{\delta}$, $\boldsymbol{\eta}$ and ϵ .

Conditioning on the first observation the likelihood of \mathbf{Y} is given by

$$L(\boldsymbol{\Theta}, \boldsymbol{\delta}, \boldsymbol{\eta}, \epsilon) = e^{-n\lambda} \prod_{t=2}^n \sum_{i=0}^{M_t} \frac{\lambda^{X_t-i}}{(X_t-i)!} C_i^{X_{t-1}} \alpha^i (1-\alpha)^{X_{t-1}-i} \quad (3)$$

with $X_t = Y_t - \eta_t \delta_t$ and $M_t = \min(X_{t-1}, X_t)$, $t = 2, \dots, n$.

The prior distribution for the contamination parameter ϵ is $\epsilon \sim \text{Be}(h, g)$, with expectation $E(\epsilon) = h/(h+g)$. Regarding the INAR(1) parameters α and λ we choose for prior distributions the conjugate of Binomial and Poisson, respectively and thus $\alpha \sim \text{Be}(a, b)$, $\lambda \sim \text{Ga}(c, d)$ (Silva et al., 2005). The set of hyperparameters a, b, c, d, β, h, g are assumed known.

Let $\pi(\boldsymbol{\Theta}, \boldsymbol{\delta}, \boldsymbol{\eta}, \epsilon)$ denote the prior distribution for $(\boldsymbol{\Theta}, \boldsymbol{\delta}, \boldsymbol{\eta}, \epsilon)$. Then

$$\pi(\boldsymbol{\Theta}, \boldsymbol{\delta}, \boldsymbol{\eta}, \epsilon) \propto e^{-d\lambda} \lambda^{c-1} \alpha^{a-1} (1-\alpha)^{b-1} \epsilon^{h-1} (1-\epsilon)^{g-1} \prod_{t=2}^n e^{-\beta} \frac{\beta^{\eta_t}}{\eta_t!} \quad (4)$$

The posterior distribution of $\boldsymbol{\Theta}$, $\boldsymbol{\delta}$, $\boldsymbol{\eta}$ and ϵ is then given by

$$\begin{aligned} \pi(\boldsymbol{\Theta}, \boldsymbol{\delta}, \boldsymbol{\eta}, \epsilon | \mathbf{y}) &\propto L(\boldsymbol{\Theta}, \boldsymbol{\delta}, \boldsymbol{\eta}, \epsilon) \pi(\boldsymbol{\Theta}, \boldsymbol{\delta}, \boldsymbol{\eta}, \epsilon) \\ &\propto e^{-[d\lambda+n\beta]} \lambda^{c-1} \alpha^{a-1} (1-\alpha)^{b-1} \epsilon^{h-1} (1-\epsilon)^{g-1} \\ &\quad \frac{\beta^{\sum_{t=2}^n \eta_t}}{\prod_{t=2}^n \eta_t!} L(\boldsymbol{\Theta}, \boldsymbol{\delta}, \boldsymbol{\eta}, \epsilon) \end{aligned} \quad (5)$$

with $0 < \alpha < 1$, $\lambda > 0$, $0 < \epsilon < 1$, and $\eta_t = 0, 1, \dots$, $t = 2, 3, \dots, n$.

The complexity of the posterior marginals of $\boldsymbol{\delta}$ and $\boldsymbol{\eta}$ suggest resorting to MCMC methods to implement the Bayesian approach described above.

The full conditional posterior distributions for α and λ are given by (Silva et al., 2005)

$$\pi(\alpha|\mathbf{Y}, \lambda, \boldsymbol{\delta}, \boldsymbol{\eta}, \epsilon) \propto \alpha^{a-1} (1-\alpha)^{b-1} \prod_{t=2}^n \sum_{i=0}^{M_t} T(t, i) \alpha^i (1-\alpha)^{X_{t-1}-i} \quad (6)$$

with $T(t, i) = \frac{\lambda^{X_{t-1}-i}}{(X_{t-1}-i)!} C_i^{X_{t-1}}$
and

$$\pi(\lambda|\mathbf{Y}_n, \alpha, \boldsymbol{\eta}, \epsilon, \boldsymbol{\delta}) \propto \lambda^{c-1} e^{-(d+n)\lambda} \prod_{t=2}^n \sum_{i=0}^{M_t} U(t, i) \lambda^{X_{t-1}-i} \quad (7)$$

with $U(t, i) = \frac{1}{(X_{t-1}-i)!} C_i^{X_{t-1}} \alpha^i (1-\alpha)^{X_{t-1}-i}$, respectively.

Now, with respect to the full conditional distribution of $\boldsymbol{\delta}$ we reason as follows. For each $j = 2, \dots, n$, $\delta_j | (\mathbf{Y}, \alpha, \lambda, \boldsymbol{\eta}, \epsilon, \boldsymbol{\delta}_{(-j)}) \sim Ber(p_j)$, where $\boldsymbol{\delta}_{(-j)}$ denotes the vector $\boldsymbol{\delta}$ with the j th component deleted. Accordingly, we can write

$$p_j = P(\delta_j = 1 | \mathbf{Y}, \alpha, \lambda, \boldsymbol{\eta}, \epsilon, \boldsymbol{\delta}_{(-j)}) = \frac{P(\delta_j = 1, \mathbf{Y} | \alpha, \lambda, \boldsymbol{\eta}, \epsilon, \boldsymbol{\delta}_{(-j)})}{f(\mathbf{Y} | \alpha, \lambda, \boldsymbol{\eta}, \epsilon, \boldsymbol{\delta}_{(-j)})} \quad (8)$$

But

$$\begin{aligned} f(\mathbf{Y} | \alpha, \lambda, \boldsymbol{\eta}, \epsilon, \boldsymbol{\delta}_{(-j)}) &= P(\delta_j = 1 | \dots) f(\mathbf{Y} | \delta_j = 1, \alpha, \lambda, \boldsymbol{\eta}, \epsilon, \boldsymbol{\delta}_{(-j)}) \\ &\quad + P(\delta_j = 0 | \dots) f(\mathbf{Y} | \delta_j = 0, \alpha, \lambda, \boldsymbol{\eta}, \epsilon, \boldsymbol{\delta}_{(-j)}) \end{aligned}$$

with $P(\delta_j = 1 | \dots) = P(\delta_j = 1 | \alpha, \lambda, \boldsymbol{\eta}, \epsilon) = \epsilon$.

Therefore

$$p_j = \frac{\epsilon f(\mathbf{Y} | \delta_j = 1, \alpha, \lambda, \boldsymbol{\eta}, \epsilon, \boldsymbol{\delta}_{(-j)})}{\epsilon f(\mathbf{Y} | \delta_j = 1, \alpha, \lambda, \boldsymbol{\eta}, \epsilon, \boldsymbol{\delta}_{(-j)}) + (1-\epsilon) f(\mathbf{Y} | \delta_j = 0, \alpha, \lambda, \boldsymbol{\eta}, \epsilon, \boldsymbol{\delta}_{(-j)})} \quad (9)$$

To compute $f(\mathbf{Y} | \delta_j = 1, \alpha, \lambda, \boldsymbol{\eta}, \epsilon, \boldsymbol{\delta}_{(-j)})$ first note that from (3) and the Markovian property of the INAR(1) model the outlier at time j affects the model for $t = j$ and $t = j + 1$. Then,

$$\begin{aligned} f(\mathbf{Y} | \delta_j = 1, \alpha, \lambda, \boldsymbol{\eta}, \epsilon, \boldsymbol{\delta}_{(-j)}) &= f(X_j, X_{j+1} | X_{j-1}, \delta_j = 1, \alpha, \lambda, \boldsymbol{\eta}, \epsilon, \boldsymbol{\delta}_{(-j)}) \\ &= f(X_j | X_{j-1}, \delta_j = 1, \alpha, \lambda, \boldsymbol{\eta}, \epsilon, \boldsymbol{\delta}_{(-j)}) \\ &\quad \times f(X_{j+1} | X_j, \delta_j = 1, \alpha, \lambda, \boldsymbol{\eta}, \epsilon, \boldsymbol{\delta}_{(-j)}) \end{aligned} \quad (10)$$

with $f(X_t | X_{t-1}) = e^{-\lambda} \sum_{i=0}^{M_t} \frac{\lambda^{X_t-i}}{(X_t-i)!} C_i^{X_{t-1}} \alpha^i (1-\alpha)^{X_{t-1}-i}$ and $M_t = \min(X_{t-1}, X_t)$ as before. Moreover, if $\delta_j = 1$ then $X_j = Y_j - \eta_j$. Therefore

$$\begin{aligned}
f(X_j|X_{j-1}, \delta_j = 1, \alpha, \lambda, \boldsymbol{\eta}, \epsilon, \boldsymbol{\delta}_{(-j)}) &= P(\alpha \circ X_{j-1} + e_j = X_j | X_{j-1}, \delta_j = 1, \dots) \\
&= P(\alpha \circ X_{j-1} + e_j = Y_j - \eta_j | X_{j-1}, \delta_j = 1, \dots) \\
&= e^{-\lambda} \sum_{i=0}^{M_j} C_i^{X_{j-1}} \alpha^i (1 - \alpha)^{X_{j-1}-i} \frac{\lambda^{Y_j - \eta_j - i}}{(Y_j - \eta_j - i)!}
\end{aligned} \tag{11}$$

and

$$\begin{aligned}
f(X_{j+1}|X_j, \delta_j = 1, \alpha, \lambda, \boldsymbol{\eta}, \epsilon, \boldsymbol{\delta}_{(-j)}) &= P(\alpha \circ X_j + e_{j+1} = X_{j+1} | X_j, \delta_j = 1, \alpha, \lambda, \eta_j, \epsilon) \\
&= e^{-\lambda} \sum_{i=0}^{M_j^*} C_i^{Y_j - \eta_j} \alpha^i (1 - \alpha)^{Y_j - \eta_j - i} \frac{\lambda^{X_{j+1} - i}}{(X_{j+1} - i)!}
\end{aligned} \tag{12}$$

with $M_t^* = \min(Y_t - \eta_t, X_{t+1})$.

Similarly, if $\delta_j = 0$ then $X_j = Y_j$ and therefore

$$\begin{aligned}
f(\mathbf{Y} | \delta_j = 0, \alpha, \lambda, \boldsymbol{\eta}, \epsilon, \boldsymbol{\delta}_{(-j)}) &= e^{-2\lambda} \\
&\prod_{t=j}^{j+1} \sum_{i=0}^{M_t} C_i^{X_{t-1}} \alpha^i (1 - \alpha)^{X_{t-1}-i} \frac{\lambda^{X_t - i}}{(X_t - i)!}
\end{aligned} \tag{13}$$

To derive the conditional posterior distribution of $\boldsymbol{\eta}$ note that if $\delta_j = 0$, no outlier at $t = j$, there is no information about η_j except the prior. Then $\eta_j | (\mathbf{Y}, \lambda, \alpha, \epsilon, \delta_j = 0, \boldsymbol{\eta}_{(-j)}) \sim Po(\beta)$. However, if $\delta_j = 1$ \mathbf{Y} contains information about η_j . Therefore,

$$\begin{aligned}
\pi(\eta_j | \mathbf{Y}, \lambda, \alpha, \epsilon, \delta_j = 1, \boldsymbol{\eta}_{(-j)}) &= \frac{\pi(\eta_j | \lambda, \alpha, \epsilon, \delta_j = 1) f(\mathbf{Y} | \lambda, \alpha, \epsilon, \delta_j = 1, \eta_j)}{\sum_{\eta_j=0}^{\infty} \pi(\eta_j | \lambda, \alpha, \epsilon, \delta_j = 1) f(\mathbf{Y} | \lambda, \alpha, \epsilon, \delta_j = 1, \eta_j)} \\
&\propto e^{-\beta} \beta^{\eta_j} / (\eta_j!) f(X_j, X_{j+1} | \eta_j, \delta_j = 1, \alpha, \lambda, \epsilon), \\
&\eta_j = 0, 1, 2, \dots
\end{aligned} \tag{14}$$

with $f(X_j, X_{j+1} | \eta_j, \delta_j = 1, \alpha, \lambda, \epsilon)$ as given in (10), (11) and (12) and $\boldsymbol{\eta}_{(-j)}$ denoting the vector $\boldsymbol{\eta}$ with the j th component deleted.

Finally, the conditional posterior distribution for ϵ depends only on $\boldsymbol{\delta}$. Since the prior distribution of ϵ is $Be(h, g)$ the conditional posterior is given by

$$\epsilon | \mathbf{Y}, \lambda, \boldsymbol{\eta}, \boldsymbol{\delta} \equiv \epsilon | \boldsymbol{\delta} \sim Be(h + k, g + n - 1 - k) \tag{15}$$

where k is the estimated number of outliers (number of δ_j 's=1).

2.2 Computational Issues

We may use the full conditional distributions of $\alpha, \lambda, \boldsymbol{\delta} = (\delta_2, \dots, \delta_n), \boldsymbol{\eta} = (\eta_2, \dots, \eta_n)$ and ϵ to draw a sample of a Markov chain which converges to the joint posterior distribution of the parameters. In most cases we can not generate directly from the full conditionals. Since they are not log-concave densities we use Gibbs methodology within Metropolis step. In particular the Adaptive Rejection Metropolis sampling - ARMS (Gilks et al., 1995) - is used inside the Gibbs sampler. When the number of iterations is sufficiently large, the Gibbs draw can be regarded as a sample from the joint posterior distribution. Accordingly there are two key issues in the successful implementation of this methodology: deciding the length of the chain and the burn-in period and establishing the convergence of the chain. We use a burn-in period of M iterations and then iterate the Gibbs sampler for a further N iterations but retain only each L th value. This thinning strategy reduces the autocorrelation within the chain.

Once the posterior probability of outlier occurrence at each time point, $p_j = P(\delta_j = 1 | \mathbf{Y}, \alpha, \lambda, \boldsymbol{\eta}, \epsilon, \delta_{(-j)})$ is estimated a cut-off point of 0.5 is used for detecting outliers, i.e. there is a possible outlier when $\hat{p}_j > 0.5$.

We now discuss the other relevant issue in the proposed bayesian approach: the choice of the hyperparameters for prior distributions. Recall from the previous section that $\alpha \sim \text{Be}(a, b)$, $\lambda \sim \text{Ga}(c, d)$. We set $a = b = c = d = 0.001$ to use non informative prior distributions (Beta and Gamma distributions with large variability). For the prior for $\epsilon \sim \text{Be}(h, g)$ we choose $h = 5$, $g = 95$ so that $E(\epsilon) = 0.05$ to reflect the prior belief that outliers occur occasionally with probability 0.05 for any time point. Regarding the parameter β of the prior distribution for the size of the outlier at time t , $\eta_t \sim P(\beta)$ two approaches are pursued: an informative setup in which β_{info} is set equal to three times the standard deviation of the 1-step-ahead prediction error and also a non-informative setup with $\beta_{info} = 30$ to reflect large variability.

3 Illustration

In this section we illustrate the performance of the above procedure with simulated data sets of 100 observations and the IP data example of section 1.

3.1 Simulated data sets

We consider time series simulated from several INAR(1) processes with $\alpha = 0.15, 0.5, 0.85$ and $\lambda = 1, 3, 5$ with one and three outliers of different sizes η of order equal to three, five and seven times the standard deviation. The times of the outliers are generated randomly.

The Gibbs sampler used to obtain the Bayesian estimates is iterated $M+N = 5000$ times and the $L = 5$ th value of the last $N = 2500$ iterations is kept, providing sample sizes of 500 values from which the estimates are computed as the sample means.

The results are reported for $\beta_{ninfo} = 30$ since they do not differ from those obtained with β_{info} .

The results are illustrated in figure 2 with simulated data from the model with parameters $\alpha = 0.85$, $\lambda = 1$, outliers at $t = 9, 29, 75$ with sizes $\eta = 7, 13, 18$, respectively. Figure 2 represents the time series and the posterior probability of outlier occurrence for each time point, \hat{p}_t . The Gibbs sampling successfully detects the outliers with estimated size of $\hat{\eta}_9 = 7$, $\hat{\eta}_{29} = 12$ and $\hat{\eta}_{75} = 19$.

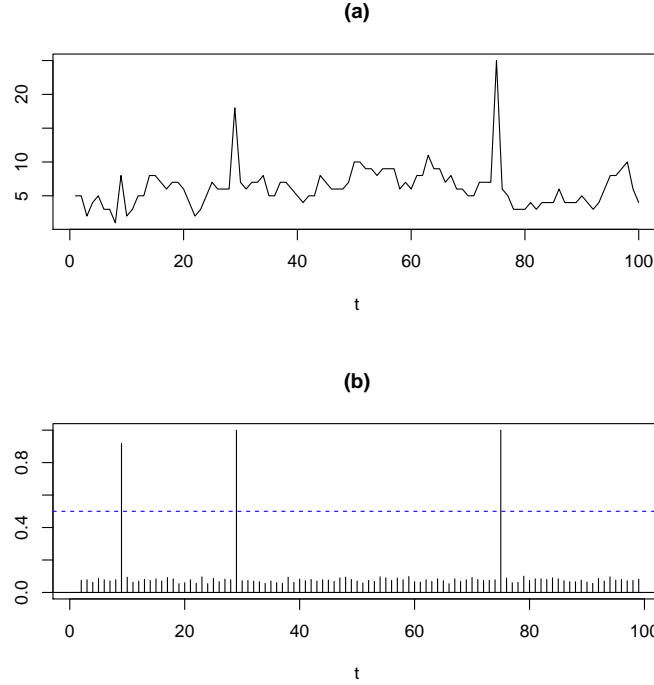


Figure 2: (a) Simulated data with $\alpha = 0.85$, $\lambda = 1$, outlier at times $t = 9, 29, 75$ with sizes $\eta = 7, 13, 18$, respectively; (b) posterior probability of outlier occurrence at each time point, $\hat{\eta} = 7, 12, 19$, respectively.

The results for all the simulated models are summarized in table 1 for series contaminated with 3 outliers. The table contains: the parameters α and λ used to generate the series with outliers of size η_S at times S , estimates for the parameters α and λ obtained by conditional least squares (assuming no outlier), *Initial CLS* and by Gibbs sampling, *Final Bayes*, the estimated probability of outlier occurrence, *Probability*, and the estimated outlier size for all the time points for which that probability is over the threshold 0.5, *Final Bayes*. For comparison purposes the table also presents the CLS estimates for the parameters α and λ after removing the effect of the detected outliers, *Final CLS*. The

results presented in table 1 indicate that the procedure is able to detect additive outliers in INAR(1) models. For models with small variability (α and λ small) small outliers are more difficult to detect. This is illustrated for an INAR(1) model with parameters $\alpha = 0.15$, $\lambda = 1$ and outliers of size $\eta_{50} = 5$ and $\eta_{34} = 7$ which are not detected. For models with larger variability the outliers are correctly detected even when their size is small (see figure 2). Moreover, the results in table 1 illustrate the negative impact of the outliers on the estimates of α and λ (*Initial CLS*). It is worthwhile noting that the estimates obtained from Gibbs sampling (*Final Gibbs*) and the conditional estimates obtained removing the effect of the detected outliers (*Final CLS*) are, in general, similar. However, for small α and for these particular simulated series, the Bayesian estimates are biased which is a typical behaviour for this range of α values (Silva et al., 2005).

3.2 IP data example

Let us consider once again the motivating example of section 1, regarding the number of different IP addresses accessing the server of the Department of Statistics of the University of Würzburg on November 29th, 2005, between 10a.m. and 6p.m., represented in figure 1 (Weiß, 2007). The sample mean and variance of the series are $\bar{x} = 1.32$, $\hat{\sigma}^2 = 1.39$. The autocorrelation and partial autocorrelation functions indicate that a model of order one is appropriate. CLS estimates for α and λ are $\hat{\alpha} = 0.22$ and $\hat{\lambda} = 1.03$, respectively. The result of applying the proposed methodology is represented in figure 3(b) indicating the possible occurrence of an outlier at time $t = 224$. The estimated size of the outlier is $\hat{\eta} = 7$. It is interesting to note that setting the time of the outlier to $t = 224$ and using the results from Barczy et al. (2011) the CLS estimate for η is $\hat{\eta}_{CLS} = 6.73$. Removing the effect of the outlier at $t = 224$ the mean and variance of the resulting series are 1.29 and 1.2, respectively. The autocorrelation and partial autocorrelation functions still indicate that a model of order one is appropriate. CLS estimates for the parameters are now $\hat{\alpha}_{CLS} = 0.29$ and $\hat{\lambda}_{CLS} = 0.91$ in accordance with the estimates obtained from the Gibbs sampling, $\hat{\alpha}_{Bayes} = 0.27$ and $\hat{\lambda}_{Bayes} = 0.89$, whose posterior distribution is represented in figure 4.

4 Concluding remarks

In this paper, the Gibbs sampling for detecting additive outliers in Poisson INAR(1) time series is presented. We estimate the probability that an observation is affected by an outlier. This procedure has the advantage of identifying observations that may require further scrutinizing. Note that the hyperparameters of the prior distributions of the outlier size and outlier occurrence probability, β and ϵ , respectively, are fixed but the same methodology applies if they are time dependent, β_t and ϵ_t , say. Masking and swamping effects caused by patches of outliers may occur depending on the size and relative position of the outliers within the patch. The solution of this problem is being investigated.

The extension of this methodology to models of higher-order, INAR(p) $p > 1$,

is possible. The mathematical expressions are easily derived from the likelihood function. However, the later and consequently the full conditional posterior distributions are highly complex. Therefore, the implementation of the methodology for higher order models requires additional computing effort. Since applications of higher-order INAR models are scarce in the literature this extension has not been considered in this work.

Acknowledgements

This work was supported by *FEDER* funds through *COMPETE* –Operational Programme Factors of Competitiveness (“Programa Operacional Factores de Competitividade”) and by Portuguese funds through the *Center for Research and Development in Mathematics and Applications* (University of Aveiro) and the Portuguese Foundation for Science and Technology (“FCT–Fundação para a Ciência e a Tecnologia”), within project PEst-C/MAT/UI4106/2011 with COMPETE number FCOMP-01-0124-FEDER-022690.

References

References

- Al-Osh, M.A., Alzaid, A.A., 1987. First-order integer-valued autoregressive (inar(1)) process. *Journal of Time Series Analysis* 8, 261–275.
- Barczy, M., Ispány, M., Pap, G., Scotto, M., Silva, M.E., 2010. Innovational outliers in inar(1) models. *Communications in Statistics - Theory and Methods* 39, 3343–3362. <http://www.tandfonline.com/doi/pdf/10.1080/03610920903259831>.
- Barczy, M., Ispány, M., Pap, G., Scotto, M., Silva, M.E., 2011. Additive outliers in inar(1) models. *Statistical Papers*, 1–1510.1007/s00362-011-0398-x.
- Chang, I., Tiao, G.C., Chen, C., 1988. Estimation of time series parameters in the presence of outliers. *Technometrics* 30, pp. 193–204.
- Chen, C., Liu, L.M., 1993. Joint estimation of model parameters and outlier effects in time series. *Journal of the American Statistical Association* 88, pp. 284–297.
- Chen, C.W., 1997. Detection of additive outliers in bilinear time series. *Computational Statistics and Data Analysis* 24, 283 – 294.
- Fokianos, K., Fried, R., 2010. Interventions in ingarch processes. *Journal of Time Series Analysis* 31, 210–225.
- Fox, A.J., 1972. Outliers in time series. *Journal of the Royal Statistical Society. Series B (Methodological)* 34, pp. 350–363.
- Gilks, W.R., Best, N.G., Tan, K.K.C., 1995. Adaptive rejection metropolis sampling within gibbs sampling. *Journal of the Royal Statistical Society. Series C (Applied Statistics)* 44, pp. 455–472.

- Justel, A., Peña, D., Tsay, R., 2001. Detection of outlier patches in autoregressive time series. *Statistica Sinica* 11, 651–673.
- McKenzie, E., 1985. Some simple models for discrete variate time series1. *JAWRA Journal of the American Water Resources Association* 21, 645–650.
- Silva, I., Silva, M., Pereira, I., Silva, N., 2005. Replicated inar(1) processes. *Methodology and Computing in Applied Probability* 7, 517–542. 10.1007/s11009-005-5006-x.
- Tsay, R.S., 1986. Time series model specification in the presence of outliers. *Journal of the American Statistical Association* 81, pp. 132–141.
- Weiß, C.H., 2007. Controlling correlated processes of poisson counts. *Quality and Reliability Engineering International* 23, 741–754.

Parameter	True	Estimates			Probability
		Initial CLS	Final		
			Bayes	CLS	
α	0.15	0.14	0.07	0.17	
λ	1	1.20	1.27	1.05	
η_{34}	7		—		0.15
η_{50}	5		—		0.07
η_{63}	9		9		0.87
α	0.15	0.09	0.01	0.03	
λ	3	3.47	3.40	3.40	
η_{34}	9		11		0.99
η_{50}	13		13		0.99
η_{63}	6		—		0.12
α	0.15	0.04	0.32	0.15	
λ	5	6.35	4.0	5.1	
η_{34}	7		—		0.09
η_{50}	12		13		0.96
η_{63}	16		18		0.99
α	0.5	0.22	0.41	0.37	
λ	1	1.04	0.94	1.05	
η_9	10		11		0.90
η_{27}	4		—		0.01
η_{97}	7		8		0.81
α	0.5	0.23	0.59	0.57	
λ	3	4.72	2.28	2.39	
η_9	17		19		0.99
η_{27}	12		16		0.99
η_{97}	7		10		0.99
α	0.5	0.26	0.51	0.57	
λ	5	7.04	4.30	3.87	
η_9	10		14		0.91
η_{27}	21		22		0.99
η_{97}	15		17		0.99
α	0.85	0.37	0.86	0.80	
λ	3	1.27	2.62	3.90	
η_9	31		29		0.92
η_{29}	13		10		0.99
η_{75}	22		22		0.99
α	0.85	0.46	0.85	0.85	
λ	5	17.55	4.60	4.66	
η_{38}	40		37		0.92
η_{41}	28		27		0.99
η_{78}	17		20		0.99

Table 1: Results from Gibbs sampling in simulated INAR(1) time series with parameters α and λ , three outliers each of size η_S at time S .

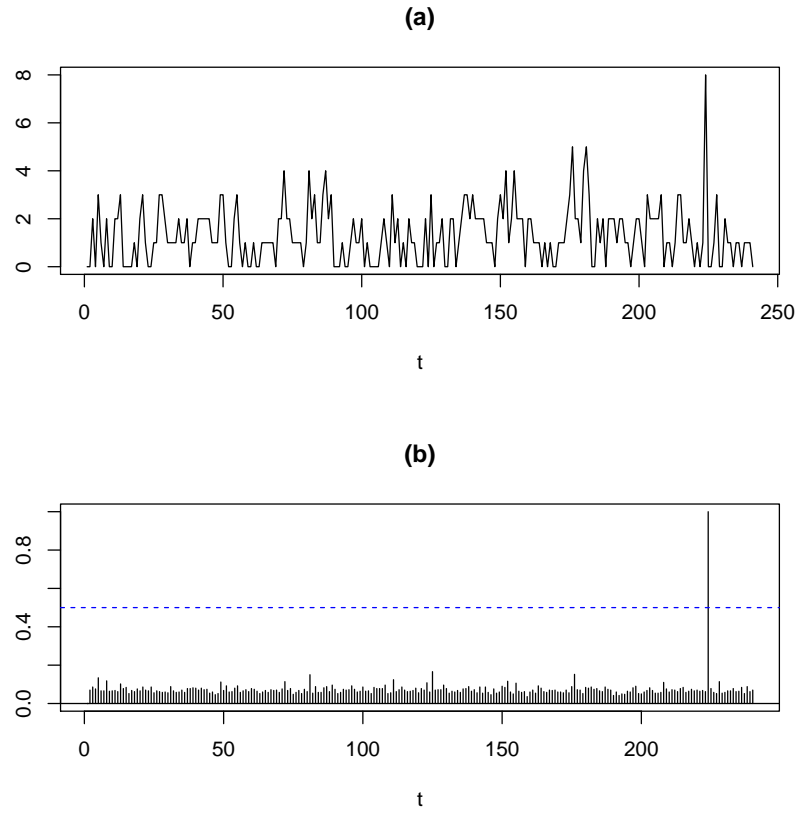


Figure 3: Number of different IP addresses accessing the server of the pages of the Department of Statistics of the University of Würzburg **(a)** and posterior probability of outlier occurrence at time t for the IP data set **(b)**.

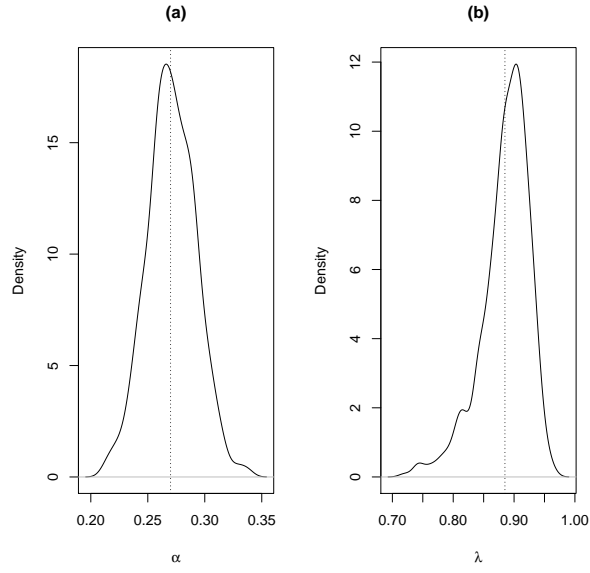


Figure 4: Posterior distribution of α and λ . The dotted lines represent the estimates $\hat{\alpha}_{Bayes} = 0.27$ and $\hat{\lambda}_{Bayes} = 0.89$.